Effect of image quality on calcification detection in digital mammography

Lucy M. Warren[a] and Alistair Mackenzie
National Co-ordinating Centre for the Physics of Mammography, Royal Surrey County Hospital NHS Foundation Trust, Guildford GU2 7XX, United Kingdom and Department of Physics, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, GU2 7XH, United Kingdom

Julie Cooke
Jarvis Breast Screening and Diagnostic Centre, Guildford GU1 1LJ, United Kingdom

Rosalind M. Given-Wilson
Department of Radiology, St. George's Healthcare NHS Trust, Tooting, London SW17 0QT, United Kingdom

Matthew G. Wallis
Cambridge Breast Unit, Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, United Kingdom and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 0QQ, United Kingdom

Dev P. Chakraborty
Department of Radiology, University of Pittsburgh, Pittsburgh, Pennsylvania 15210

David R. Dance
National Co-ordinating Centre for the Physics of Mammography, Royal Surrey County Hospital NHS Foundation Trust, Guildford GU2 7XX, United Kingdom and Department of Physics, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford GU2 7XH, United Kingdom

Hilde Bosmans
University Hospitals Leuven, Herestraat 49, 3000 Leuven, Belgium

Kenneth C. Young
National Co-ordinating Centre for the Physics of Mammography, Royal Surrey County Hospital NHS Foundation Trust, Guildford GU2 7XX, United Kingdom and Department of Physics, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford GU2 7XH, United Kingdom

(Received 31 January 2012; revised 12 April 2012; accepted for publication 27 April 2012; published 17 May 2012)

Purpose: This study aims to investigate if microcalcification detection varies significantly when mammographic images are acquired using different image qualities, including: different detectors, dose levels, and different image processing algorithms. An additional aim was to determine how the standard European method of measuring image quality using threshold gold thickness measured with a CDMAM phantom and the associated limits in current EU guidelines relate to calcification detection.
Methods: One hundred and sixty two normal breast images were acquired on an amorphous selenium direct digital (DR) system. Microcalcification clusters extracted from magnified images of slices of mastectomies were electronically inserted into half of the images. The calcification clusters had a subtle appearance. All images were adjusted using a validated mathematical method to simulate the appearance of images from a computed radiography (CR) imaging system at the same dose, from both systems at half this dose, and from the DR system at quarter this dose. The original 162 images were processed with both Hologic and Agfa (Musica-2) image processing. All other image qualities were processed with Agfa (Musica-2) image processing only. Seven experienced observers marked and rated any identified suspicious regions. Free response operating characteristic (FROC) and ROC analyses were performed on the data. The lesion sensitivity at a nonlesion localization fraction (NLF) of 0.1 was also calculated. Images of the CDMAM mammographic test phantom were acquired using the automatic setting on the DR system. These images were modified to the additional image qualities used in the observer study. The images were analyzed using automated software. In order to assess the relationship between threshold gold thickness and calcification detection a power law was fitted to the data.
Results: There was a significant reduction in calcification detection using CR compared with DR: the alternative FROC (AFROC) area decreased from 0.84 to 0.63 and the ROC area decreased from 0.91 to 0.79 ($p < 0.0001$). This corresponded to a 30% drop in lesion sensitivity at a NLF equal to 0.1. Detection was also sensitive to the dose used. There was no significant difference in detection between the two image processing algorithms used ($p > 0.05$). It was additionally found that lower threshold gold thickness from CDMAM analysis implied better cluster detection. The measured threshold gold thickness passed the acceptable limit set in the EU standards for all image qualities except half dose CR. However, calcification detection varied significantly between image qualities. This suggests that the current EU guidelines may need revising.

 0094-2405/2012/39(6)/3202/12/$30.00

Conclusions: Microcalcification detection was found to be sensitive to detector and dose used. Standard measurements of image quality were a good predictor of microcalcification cluster detection.
[http://dx.doi.org/10.1118/1.4718571]

Key words: digital mammography, dose, image processing, observer performance, calcification

I. INTRODUCTION

Commercially available direct digital (DR) and computerized radiography (CR) mammography systems vary greatly in terms of physical performance and cost. Some large clinical trials have provided insight on the clinical performance of various detectors.[1,2] One of these, the ACRIN DMIST study[1] was a large prospective study in which patients were imaged on both screen-film and digital systems. This study found the overall diagnostic accuracy for screen-film and digital to be similar, but digital mammography was more accurate in women under the age of 50 yr, women with radiographically dense breasts, and pre- or perimenopausal women.

A subsequent retrospective study compared accuracy for cancer diagnosis of separate digital mammography manufacturers with screen-film mammography.[3] This study found that there was no significant difference in accuracy of cancer diagnosis for soft-copy digital and screen-film mammography for the three digital manufacturers investigated. The authors did not intend to compare one digital manufacturer's equipment to another but to compare digital with screen-film mammography for each digital manufacturer. Since different women were imaged on different digital systems it would not have been possible to perform a paired comparison between different manufacturers.

As well as different digital systems, each system can be operated in a variety of ways that affects its performance. The automatic exposure control (AEC) can be set to give a higher or lower dose for a particular breast thickness. Also each manufacturer's system uses a different image processing algorithm. Therefore, as well as investigating the effect of different detectors on calcification detection, it is also important to investigate the effect of dose and image processing.

Previous work comparing the detection of microcalcification clusters in mammographic images have found mixed results.[4,5] A phantom study[4] found that an a-Se FP detector outperformed a CCD system, a high resolution CR detector, and a SF system. Finally, a prospective clinical study[5] found that detection of lesions containing calcification was higher with an a-Si DR detector compared to CR. Both these studies, however, suffered from one of the following limitations: images of phantoms were used rather than clinical images,[4] hard copy of digital images was viewed[4,5] (this is not current clinical practice in the United Kingdom), subjective measures of image quality were used[5] and sophisticated ROC (response operating characteristic)/Free ROC analysis was not performed.

The effect of change in dose on microcalcification detection has also been investigated in the past.[6–8] Two studies using phantom and clinical images with simulated calcifications found a significant reduction in detection at quarter the dose level but not half the dose level.[7,8] Both these studies used a location-known exactly experimental paradigm. Using this paradigm is a limitation if it is aimed to predict the clinical performance that requires both search and classification.[9] A further study[6] inserted simulated mass and calcification clusters into clinical images. These images were then simulated to 50% and 30% of the original dose level. The entire image was viewed during the study and observers were required to localize the lesions prior to rating the lesion on how confident they were that they had identified a lesion. This study found a significant reduction in detection at 50% and 30% the original dose level.

Finally, several previous studies have investigated the effect of image processing on calcification detection.[10–13] One small study[12] found no significant difference in detection due to image processing. Three other studies[3,10,13] compared digital images with different image processing to screen-film imaging, but did not compare image processing software directly. The remaining study by Zanca *et al.*[11] used unprocessed clinical images into which simulated calcification clusters were inserted. This study found significant differences in detection between several image processing algorithms.

In theory, one could investigate calcification detection by organizing clinical trials to measure detection for different image qualities. In practice, such trials would be very expensive due to the low incidence of cancer in screened populations and there would be ethical issues if there were multiple radiation exposures on the same women. The alternative approach used here mimics mammographic imaging by inserting simulated calcification clusters into normal digital mammograms acquired with an amorphous selenium DR detector. The images were then modified using a method described previously[14] to have the appearance of images acquired on a CR detector typical of the single-sided granular type widely used in Europe. The images were also modified to have the appearance as if they were acquired on both systems at lower dose levels. An advantage of this approach is that the inserted calcifications and anatomical backgrounds were identical between the different test sets used in the observer study. This is important because variations in anatomical background are expected to have a major impact on cancer detection and are a confounding factor in studies of this type. Thus, our approach allowed paired comparisons of images of different image qualities and was suitable for FROC analysis, overcoming the limitation of some of the previous studies described above.

There are various methods of measuring the physical performance of digital mammographic imaging systems. Currently acceptable and achievable limits of image quality as set in the European protocol for the quality control of the physical and technical aspects of mammography screening[15] are based on images obtained with the CDMAM mammography test object (Artinis Medical Systems, Zetten, The Netherlands). The acceptable limit is the border between acceptable and unacceptable performance of a system and the achievable limit is the performance which systems should aim to achieve. However, it is uncertain how the results of such tests relate to clinical performance. An obvious criticism of the test object is that it lacks structured noise equivalent to normal breast structure. Another difference is that current measures of image quality use unprocessed images of test objects, whereas radiologists only look at images that have undergone additional processing, e.g., by reducing noise and/or enhancing edges and contrast. There has been very little work investigating this in the past. Carton *et al.*[16] performed three different contrast resolution tests for four different detectors: noise equivalent quanta (NEQ), contrast-detail curves using the CDMAM phantom and detection of simulated calcifications in mastectomy samples. All three tests ranked the systems in the same order but the correlation between contrast-detail curves and detection of simulated calcification was not directly reported. There is an urgent need to assess how performance assessed using the CDMAM phantom relates to cancer detection and whether the current limits for image quality are set appropriately.

In summary, this work aims to answer three questions: Is calcification detection affected by a change in digital mammography detector, dose level, and image processing? How do measures of the physical performance of digital mammography systems relate to calcification detection? Are the current limits for such physical measurements set appropriately?

II. MATERIALS AND METHODS

The methodology comprises the following parts (a) case preparation for observer study (b) observer study, and (c) CDMAM measurements. The study protocol was approved by the regional research ethics committee.

II.A. Case preparation

The case preparation consisted of the stages (1) breast image collection, (2) cluster collection, (3) cluster characterization, (4) image creation, (5) image modification, and (6) image processing.

II.A.1. Breast image collection

One hundred and sixty two unprocessed anonymous patient images were collected from a Hologic Selenia mammography unit (Hologic, Inc., Bedford, MA). These images were read as normal but there has been no follow up to confirm this. Each image used in the study was a single view of one patient's breast: the view (CC or MLO) and breast (left or right) were randomly selected. Ideally, two views of both breasts would be displayed as would be the case in the clinic. However, only 2D simulated calcification cluster images were available and these could not be reoriented for insertion into both views.

Patients were aged between 47 and 73 yr and had been referred for a mammography examination as they were either symptomatic or high risk. The compressed breast thicknesses ranged from 24 to 92 mm (average of 54 mm), mean glandular dose for a 50–60 mm breast was 2.09 mGy, calculated using the methodology described in the European protocol.[15] The images were acquired at 24–34 kVp (average of 29 kVp), with a target/filter combination of either molybdenum/molybdenum or molybdenum/rhodium. The pixel size was 70 μm.

II.A.2. Cluster collection

Simulated calcification clusters were extracted from unprocessed digital images of sliced mastectomy samples (average thickness 5 mm) from seven patients acquired at $\times 5$ magnification (effective pixel size of 10 μm) on a digital x-ray specimen cabinet (Model No. MX 20 DC2 Faxitron Bioptics, LLC, Lincolnshire, IL). All clusters used were from malignant lesions. The method used for calcification cluster simulation was a slightly modified version of that used previously, which is described extensively.[11,17,18] Therefore, this text concentrates mainly on the differences of this method from that used previously. In the work of Zanca *et al.*,[11] the entire calcification clusters were extracted in a single operation. This has the advantage of maintaining the spatial orientation of the calcifications within the cluster, which can be a predictive factor of malignancy. However, the anatomical background within the cluster can generate undesirable structures surrounding the calcifications. Therefore, in this work, each calcification was extracted separately. A small region of interest (ROI_{calc}) containing each calcification was cropped from the image of the breast tissue sample. The mean background pixel value within ROI_{calc} was calculated from the mean pixel value of smaller regions drawn within ROI_{calc} but outside of the calcification. Each pixel value in ROI_{calc} was divided by the mean background pixel value, generating a normalized calcification image. The pixels in ROI_{calc} with values higher than the mean background (due to the nonuniform soft tissue structure) resulted in pixel values greater than one in the normalized calcification image. These were assigned a value of one, so that when the calcification cluster was multiplied by the unprocessed breast image during insertion, these pixels associated with the background were not altered in the unprocessed breast image. This may have left background in the vicinity of the calcification due to background pixel values being lower than the mean. This method does, however, avoid the creation of unrealistically sharp edges on the extracted calcifications.

This was repeated for all calcifications within the cluster and the calcifications were recombined into the original spatial configuration creating a normalized cluster image.

Thirty-three unique normalized cluster images were generated from seven mastectomy slices in this way (several clusters were extracted from each sample). At ×5 magnification, the field of view of the specimen cabinet was 1 cm × 1 cm and so all clusters encompassed an area smaller than 1 cm × 1 cm.

The number of unique cluster templates which could be generated was limited by the number of mastectomy samples available during the period of preparation for the study. However, it was preferable to use a large number of clusters in the study as this is expected to increase the statistical power.[19] Therefore, to increase the numbers available the calcifications within each cluster were rearranged. When rearranging the calcifications the orientations of the calcifications were unchanged but the positions of their centroids were swapped. The rearranged clusters were inspected by a radiologist who confirmed that they looked sufficiently different in the image to the original clusters, and so could be treated as unique clusters. Including the original 33 clusters, this gave a total of 66 clusters. Forty-seven of the clusters were repeated twice within the image set; however, each repeated cluster was rotated and/or flipped horizontally or vertically in relation to the original cluster. The final number of clusters inserted into the clinical images was 113.

II.A.3. Cluster image characterization

The area of each calcification was characterized by calculating the number of pixels covered by a calcification. This was then converted to equivalent area (square millimeter). From this, the calcification diameter was defined as the diameter of a disc with this area.

Each cluster was quantified in terms of its diameter and the number of contained calcifications. Each cluster contained calcification in different spatial configurations. Therefore, to compare the sizes of clusters, the area was calculated using the convex hull, i.e., the smallest polygon which would fully enclose the cluster. The cluster size was then calculated as the diameter of a disc of area equal to that calculated using the convex hull.

II.A.4. Cluster image contrast and sharpness correction

As described in Sec. II.A.2, each cluster was imaged in a digital specimen cabinet at high magnification within a thin specimen. If the cluster had been within a patient's breast when imaged its appearance would have differed because of the decreased magnification, increased scatter and decreased subject contrast due to increased thickness of surrounding breast tissue and the different imaging system and imaging parameters used when acquiring the image. Therefore, prior to inserting the cluster template into a normal breast image, the appearance of the cluster image was adjusted to account for these changes. First, the contrast of the calcification cluster was adjusted appropriately using aluminum and poly(-methyl methacrylate) (PMMA) as calibrating materials which were imaged in the x-ray specimen cabinet and on the DR detector using the appropriate imaging parameters. The suitability of aluminum as a material to represent calcifications has been demonstrated previously.[20,21] Accordingly, in our measurements, aluminum was imaged on top of PMMA of thickness equivalent to the sliced mastectomy samples and the compressed breast thickness of the clinical image into which the simulated cluster will be inserted. Therefore, the effect of scatter was incorporated in the measurement. The equivalence described in Dance *et al.*[22] and also used in the European protocol[15] between different thicknesses of PMMA and breasts of typical composition was assumed. In effect, this procedure takes account of the breast thickness and assumes typical composition. It does not account for the actual composition of the individual breast corresponding to each mammogram. However, the error involved is small and the accuracy adequate for this purpose.[23–25] The convertor blur and focal spot blur of the specimen cabinet x-ray was then effectively removed and the convertor blur and focal spot blur of the DR detector applied using the modulation transfer function (MTF) of both systems, thereby correcting the sharpness of the cluster image. Finally, the cluster image was resampled using nearest neighbor averaging at the pixel size of the DR detector. This process has been described and validated elsewhere.[17,18,26,27]

II.A.5. Image creation

Half of the 162 normal breast images collected (see Sec. II.A.1) had 1–3 cluster templates inserted [using the software SARA (Ref. 28)]. Of the 81 abnormal images 70% had one cluster inserted, 20% had two clusters inserted and 10% had three clusters inserted (replicating the proportions used by Zanca *et al.*[11]). This encourages readers to search for more clusters after finding one or two and replicates clinical reality. At the location of each cluster insertion, the pixel value of each pixel in the normalized cluster image was multiplied by the pixel value of the corresponding pixel in the unprocessed linearized breast image. The regions of breast tissue into which the clusters were inserted were assigned to the broad categories: glandular, fatty, mixed with high density structure, and mixed without high density structure. The regions were selected so that there were equal numbers of each category in the study.

Clusters were inserted in any location over the breast (excluding the nipple and skin edge). All images were reviewed by an experienced radiologist for the realism of the cluster and its location in the breast. This resulted in 81 images with inserted calcification clusters (abnormal images) and 81 images which remained normal. These 162 images are referred to below as "normal dose DR" images.

Examples of calcifications inserted into each of the four categories of background (taken from four different clinical images) are shown in Fig. 1.

II.A.6. Image modification

The normal dose DR images were copied and modified to generate image sets at four different image qualities: CR images at the same dose level as the normal dose DR images ("normal dose CR") and at half this dose level ("half dose

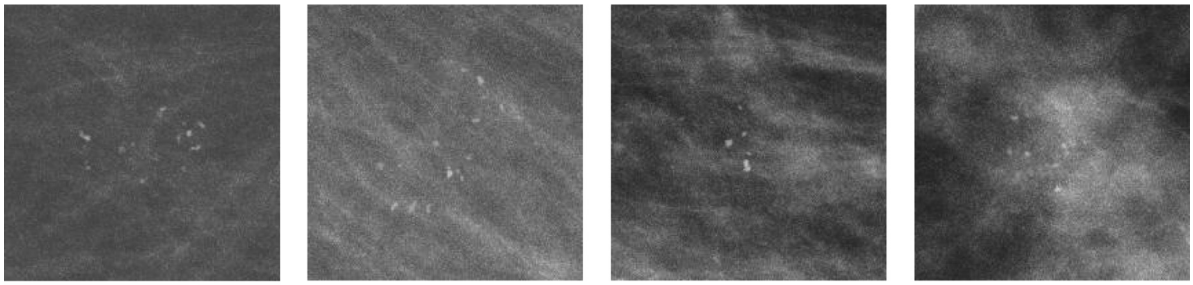

FIG. 1. Examples of clusters used in the observer study inserted in four different categories of background: (a) fatty, (b) mixed without high density structure, (c) mixed with high density structure, and (d) glandular. Each image segment is 200 × 200 pixels (pixel size of 70 μm.

CR"), and DR images at half and quarter the original dose level ("half dose DR" and "quarter dose DR," respectively). The CR system simulated is not exactly like one particular CR system but instead representative of the so called "phosphor powder plates" available. We have characterized the system which has been simulated in terms of performance with the CDMAM test object to investigate how its performance compares with other systems. This simulation was achieved using the validated methodology described by Mackenzie *et al.*[14] where differences in the MTF, noise power spectra, and signal transfer properties were accounted for.

The modification process to simulate the four different image qualities described above consists of several stages. In the description of these stages, the DR system is called the original system and the image quality being simulated, e.g., half dose DR or normal dose CR, is called the target system. First, the unprocessed image from the original system was linearized. This linearized image was then blurred in the Fourier domain by multiplying by the ratio of the presampled MTF of the target system to the presampled MTF of the original system. The blurred image was in turn then multiplied by a factor to account for the dose change between the target and original systems. The latter multiplication does not account correctly for the change in image noise between the original and target systems and a zero mean image with the appropriate amount of correlated noise was finally added to account for this difference. The simulation procedure required the NEQ of the image being simulated to be poorer than the NEQ of the image from which it is being simulated. This holds true for CR and DR at the same dose level as shown in Fig. 2(a) which displays the NEQ of the Hologic Selenia DR and simulated CR at an incident air kerma to the detector of 89 μGy. The simulated CR image was not resampled and so had a pixel size of 70 μm. The simulation methodology has been validated using quantitative measurements of the MTF, NPS, and threshold gold thickness (using the CDMAM test object) for real and simulated images.[14] The MTF and NNPS of the simulated images closely matched that of the real images. The MTFs of the DR and simulated CR system are shown in Fig. 2(b). The threshold gold thicknesses for the CDMAM test object for the simulated and real images were also similar.

II.A.7. Image processing

The final step in the imaging chain prior to display was processing the images using a commercial image processing package. The two packages available were from Hologic who manufactured the DR system used to obtain the original images and Agfa Musica-2 (Agfa Healthcare, Mortsel, Belgium). The normal dose DR images were processed with

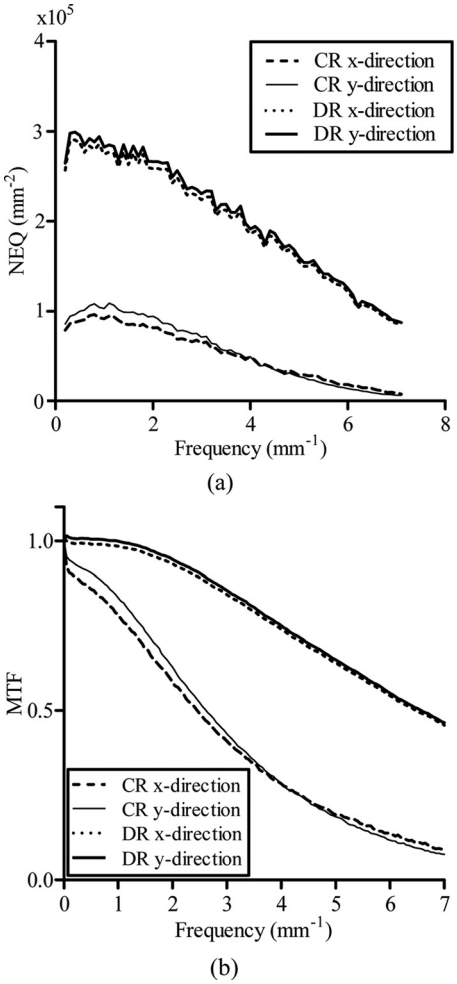

FIG. 2. Characterization measurements on Hologic Selenia DR system and simulated CR used in study. (a) NEQ for DR and simulated CR (incident air kerma to detector = 89 μGy) and (b) MTF for DR and simulated CR.

both Hologic and Agfa image processing software. All other image quality levels were processed only with the MUSICA software. The MUSICA software is useable for a range of detectors and has been routinely used with images from systems using both DR and CR detectors making it suitable for this study where two detector types were being compared. Examples of the same portion of an image with inserted cluster, at the six different image qualities described in Sec. II.A, are shown in Figs. 3(a)–3(f). In the study, observers inspected the entire image not just the portion shown in the figure.

II.B. Study protocol

Seven observers took part in the study: five radiologists and two radiographers all certified as defined by the UK NHS Breast Screening Programme[29] and reading a minimum of 5000 examinations per year. Each observer read 972 images (162 images at each of six image qualities). The observer study was performed in our laboratory. The room illuminance was monitored before each session and was maintained at the ambient level of 3.6–8.2 lx.[30] The breast images were displayed on calibrated 5 megapixel monitors (BARCO Model: MDMG-5121) using SARA software[28] and were scaled to fit the monitors. Since the DR and CR images used in the study had the same pixel size this scaling factor was the same for all image qualities. After discussion with local radiologists, this method of displaying images mirrored clinical practice. An electronic magnification tool (dimensions 200×200 pixels) was provided which could be moved freely over the image to display the selected portion of the image so that one monitor pixel corresponded to one image pixel. The observers were asked not to alter the window or zoom the entire image.

Every observer inspected 162 images in a session with a break after 81 images. To minimize the effect of learning during the course of the study, equal numbers of images from the six image qualities were randomly distributed in each session with a minimum of 1 week between sessions. Prior to the study observers were trained for the task using a pilot study of 60 images and 50 calcification clusters not included in the current study. Before starting each session, the observers reviewed a short training set comprising five images not included in the study to refamiliarize with the task and the software. The observers were asked to ignore single calcifications and vascular calcification. On identifying a suspicious region within an image, the observer marked its location and assigned a rating on a five point scale according to their confidence that the suspicious region was a calcification cluster:

1. Probably not a calcification cluster
2. Possibly a calcification cluster
3. Somewhat confident this is a calcification cluster
4. Moderately confident this is a calcification cluster
5. Very confident this is a calcification cluster.

Observers were encouraged to use the full range of scores as required for statistical analysis. The co-ordinates of the marked location and the rating were recorded automatically for subsequent statistical analysis.

II.C. Statistical analysis

The most commonly used method for comparing the performance of different imaging modalities is ROC analysis.

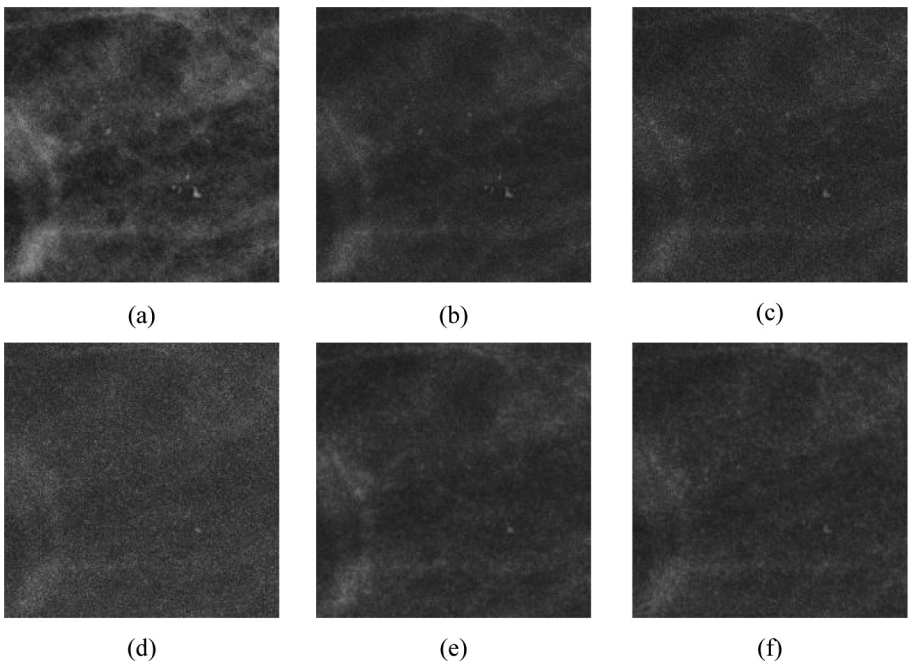

FIG. 3. Processed regions of interest (200×200 pixels) from an image in the study at all image qualities; (a) normal dose DR images with Hologic image processing, (b) normal dose DR images with Agfa image processing, (c) half dose DR with Agfa image processing, (d) quarter dose DR images Agfa image processing, (e) normal dose CR with Agfa image processing, and (f) half dose CR with Agfa image processing.

However, ROC analysis does not take into account the location of the disease or multiple disease sites on a single patient image. FROC analysis can be used to overcome this limitation and has been used in similar studies.[6,11,31,32] When inspecting an image, if the observer correctly identifies the location of a cancer within a specified area, this results in a lesion localization (LL). If the observer marks outside this specified area or marks a region on a normal image, this is a nonlesion localization (NL). In this study, the specified area was the area within the smallest rectangle bounding the cluster. The points on the alternative FROC (AFROC) curve are the lesion localization fractions (LLFs) against the fraction of normal images with at least one NL at a particular level of confidence. The most lax point is when all suspicious regions marked within the rectangle containing the cluster with a score 1–5 are considered a LL. The most stringent point is when only suspicious regions marked within the specified area with a score 5 are considered a LL. There are then three intermediate points between these two extremes. The performance of an observer inspecting images of a particular image quality is expressed as a figure of merit (FoM) calculated as the trapezoidal area underneath the points in the AFROC curve when joined with a line segment to (1, 1).[33] In this study jack-knifed FROC (JAFROC)[34] was performed (using the software JAFROC Version 4.0). Using this methodology, each case was removed in turn from the analysis, the FoM recalculated and a pseudovalue calculated using the following equation:[34]

$$PV_{ijk} = N_T FoM_{ijk} - (N_T - 1) FoM_{ij(k)}. \quad (1)$$

Here, FoM_{ijk} is the figure of merit for the ith modality and the jth reader when all the cases are used in the calculation, and $\mathrm{FoM}_{ij(k)}$ is the figure of merit for the ith modality and the jth reader when case k is deleted. N_T is the total number of cases.

The pseudovalues were then analyzed using the Dorfman-Berbaum-Metz (DBM) analysis of variance (ANOVA) technique,[35] and the reader-averaged JAFROC FoM for each image quality and the associated 95% confidence intervals (CIs) were calculated. The confidence limits for differences in FoM between different image qualities were estimated using the same methodology.

Since ROC analysis is a common method of analysis for studies comparing modalities, this was also performed as a secondary analysis. The equivalent ROC rating was assumed to be the highest FROC rating for a given image and a score of 0 was assigned to any unmarked images. ROC analysis was performed using the software DBM MRMC Beta Version 2.0. The figure of merit for ROC analysis is the area underneath the trapezoidal ROC curve.

Finally, many previous studies using FROC analysis quote lesion localization fraction (LLF) and nonlesion localization fraction (NLF) pairs.[6,31,36,37] We have also calculated the LLF at a NLF of 0.1. The NLF rate seen clinically would be much lower; however, this was the lowest possible NLF value which allowed stable results.

II.D. CDMAM measurements

Images of a CDMAM test object were obtained as described in the European protocol.[15] The CDMAM test object (Fig. 4) consists of a matrix of gold discs with thicknessses from 0.03 to 2 μm and diameters from 0.06 to 2 mm on a 2 mm aluminum base encased in PMMA.

The phantom was positioned with 20 mm PMMA blocks above and below. Sixteen images of the CDMAM phantom were acquired using the AEC selected imaging parameters (29 kV molybdenum target and rhodium filter) on the DR system. The CDMAM images were then modified as described in Sec. II.A.6 to have the appearance of CR images at the same and half the original dose level and DR images at half and quarter the original dose level. Each set of unprocessed images was automatically analyzed to determine the threshold gold thickness at all disc diameters.[38,39] The measurements were then compared to the limiting threshold gold thickness for the 0.1 and 0.25 mm gold disc diameters as set out in the European protocol.[15]

III. RESULTS

III.A. Calcification characterization

The diameter of the clusters used in the study ranged from 1.23 to 8.50 mm with a mean value of 3.89 mm [Fig. 5(a)] and the number of calcifications per cluster ranged from 7 to 46 with a median value of 16 [Fig. 5(b)]. The diameter of a single calcification ranged from 0.05 to 1.18 mm with a mean value of 0.21 mm.

III.B. Observer study

The raw data points and fitted AFROC curves for two observers who took part in the observer study are shown in Fig. 6 (the two observers were randomly selected from the total seven). Error bars are shown for three image qualities and display the 95% confidence interval calculated by bootstrapping the data. The reader-averaged AFROC curve for all six image qualities is shown in Fig. 7.

The figures show that there is a large drop in performance both with decreasing dose and also when inspecting CR

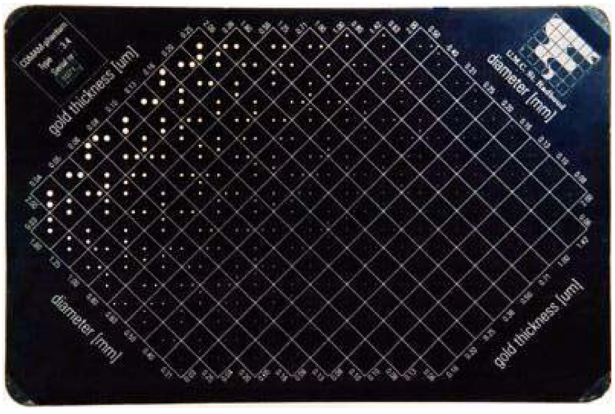

FIG. 4. CDMAM test object.

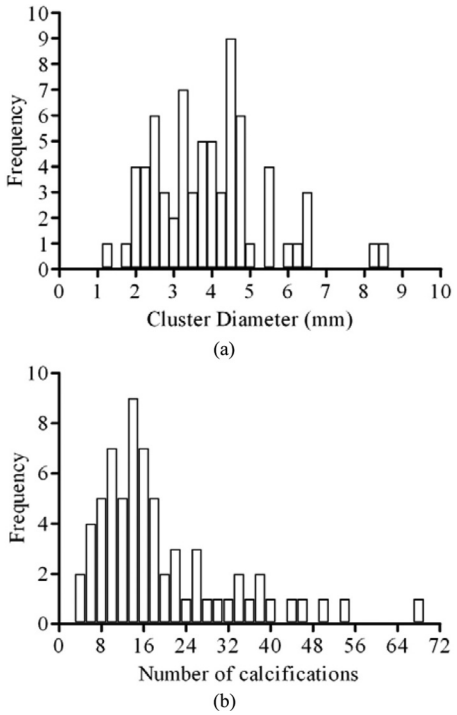

FIG. 5. Characteristics of extracted calcification clusters: (a) Cluster diameters and (b) number of calcifications in a cluster.

images compared to DR. The reader-averaged FoM and 95% CI for both JAFROC and ROC analysis are given in Table I.

The reader-averaged LLF (or lesion sensitivity) at NLF equal to 0.1 is also given Table I. Lesion sensitivity dropped by 30% at this NLF for CR compared to DR at the same dose level. There were also 25% and 12% drops in lesion sensitivity when halving the dose for DR and CR, respectively.

The ANOVA procedure displayed significant differences in performance between several image quality pairs (Fig. 8). The error bars are the associated 95% confidence limits. The difference in FoM between two image qualities is significant if the 95% confidence interval does not include zero. From this figure, it can thus be seen that the same image quality pairs were found to be significantly different from one another when using ROC or JAFROC analysis. The other eight possible pairs compare image qualities where multiple factors have changed, e.g., normal dose DR versus half dose CR and so are less noteworthy. The observers' detection performance decreased significantly when inspecting CR images compared to DR images at the same dose level with Agfa image processing (change in JAFROC FoM of −0.20 [95% confidence interval CI: −0.25, −0.14]). Detection performance also decreased significantly when halving the DR and CR doses (change in JAFROC FoM of −0.15 [95% CI: −0.21, −0.10] and −0.09 [95% CI: −0.14, −0.03] for DR and CR, respectively). In addition, there was a significant difference in detection performance between DR at half and quarter the dose level (change in JAFROC FoM of −0.15 [95% CI: −0.21, −0.10]). Finally, the observers' performance did not differ significantly between the two types of

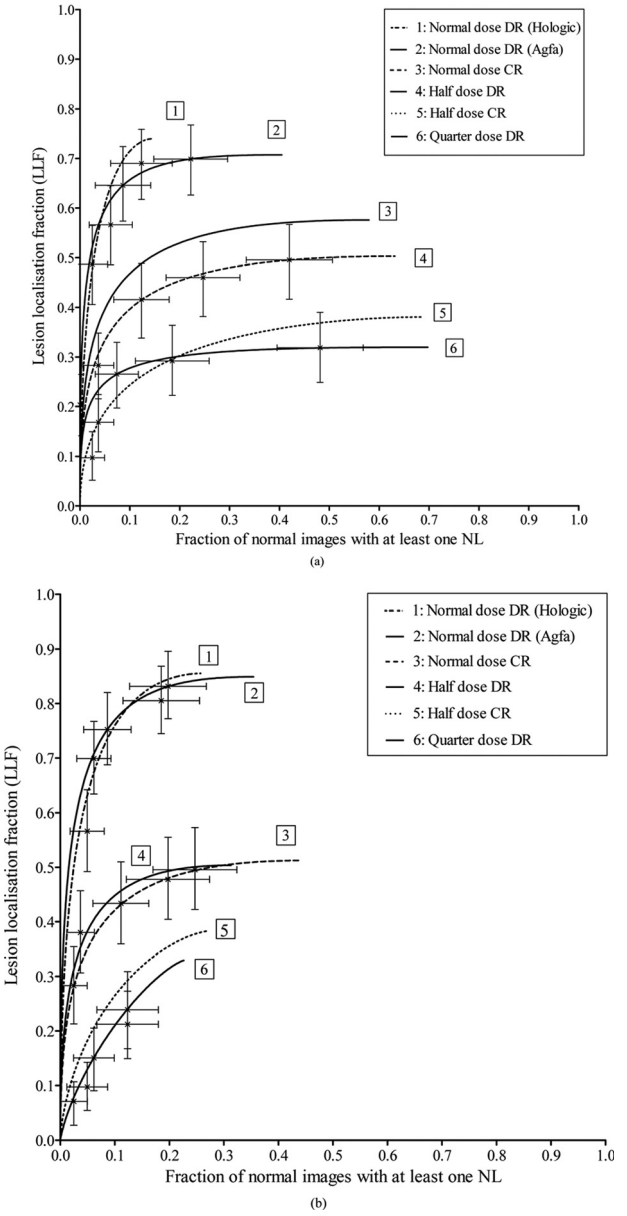

FIG. 6. AFROC curves for two observers [(a) and (b)] inspecting the six image qualities. The crosses show the raw data points calculated using the confidence scores for three of the image qualities. The associated error bars show the 95% confidence intervals estimated using bootstrapping. Data points and error bars for the remaining three image qualities have been omitted to reduce clutter in the figure.

image processing investigated (change in JAFROC FoM of −0.005 [95% CI: −0.060, 0.05]).

III.C. CDMAM measurements

The threshold gold thickness for each image quality is plotted against the corresponding dose for the 0.1 and 0.25 mm disc diameters in Figs. 9(a) and 9(b), respectively. A power law has been fitted to the data as performed when test-typing digital mammography equipment.[15,40] The acceptable and achievable limits for the 0.1 and 0.25 mm disc diameters set in the European protocol[15] are also shown, along with the dose limit for a breast equivalent to 50 mm

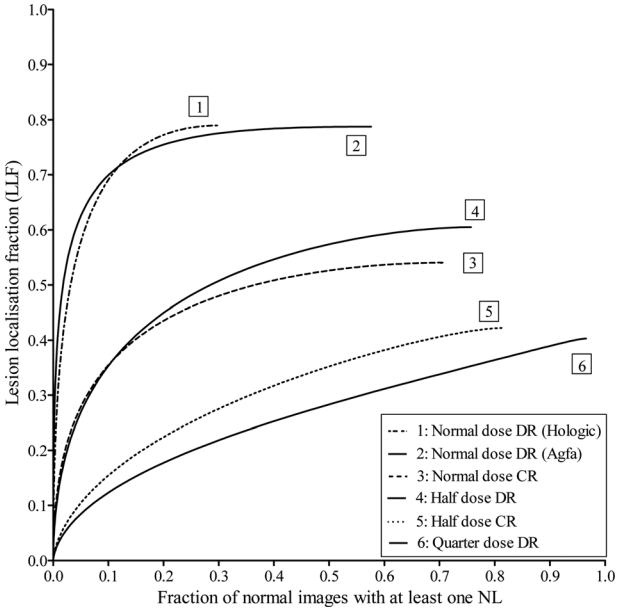

FIG. 7. Reader-averaged AFROC curves showing performance at all six image qualities.

thickness of PMMA. Following the European protocol, a system must perform below the dashed line denoting the acceptable limit and to the left of dashed line denoting the dose limit. From the figures, it is evident that a wide range in performance was simulated in this study. However, all image qualities apart from half dose CR passed the acceptable limit.

III.D. Relating CDMAM threshold gold thicknesses and observer performance

In order to determine the relationship between CDMAM threshold gold thickness and reader performance, the reader-averaged JAFROC figure of merit for each image quality has been plotted against the threshold gold thickness determined from the CDMAM images for the 0.1 and 0.25 mm disc diameters [Figs. 10(a) and 10(b)]. A power law relationship has been fitted to the data, which shows a strong correlation for both detail sizes ($R^2 = 0.94$ and $R^2 = 0.88$ for 0.1 mm and 0.25 mm disc diameters, respectively). The probability that such a relationship would occur by chance is less than or equal to 5% ($p \leq 0.05$). It can therefore be concluded that lower threshold gold thickness suggests good calcification cluster detection, and so CDMAM measurements provide a measure of performance which is relevant to calcification detection.

IV. DISCUSSION

This observer study found significant differences in calcification detection between several of the image quality pairs investigated. There was a significant reduction in calcification detection when inspecting CR images compared with DR images. This result is consistent with the prospective study by Schueller *et al.*[5] in which 150 patients were imaged on both a CR and DR detector. More calcifications were detected using DR compared with CR. We feel our work overcomes several limitations of this study. In the study by Schueller *et al.*,[5] images were printed on film before being read, whereas our study better replicated clinical practice in the United Kingdom where the majority of digital images are read soft copy. Additionally, when imaging the same patient on two different modalities (CR and DR), it is impossible to achieve exactly the same patient position and compression. This would result in a difference in the appearance of the breast structure in the images, which would have an unmeasureable effect on cancer detection. In our study, however, the use of simulated images allows a paired comparison of CR and DR with no difference in positioning. The simulated CR image quality levels were representative of the so called "powder phosphor" CR plates, as opposed to the needle photostimulable phosphor based CR systems which have been shown to have better physical performance.[41]

Microcalcification detection was also found to be significantly reduced when halving and quartering the dose level. In two previous studies[7,8] a significant difference in calcification detection was evident between full and quarter dose images but not between full and half dose images. In both these studies, the localization aspect of the task was removed. The result found in our study is, however, in agreement with the previous study by Ruschin *et al.*[6] which had a very similar design to the study described in this current work. There are, however, some differences between the two studies. Ruschin *et al.* used a smaller number of unique clusters (5 compared to 66 in the present study) and the cluster sizes were much larger (average diameter = 9.1 mm compared to 3.9 mm in this study). Finally, Ruschin *et al.* only

TABLE I. Reader-averaged JAFROC and ROC FoM and reader-averaged LLF at a NLF equal to 0.1 for all six image qualities. Image qualities 1 and 2 are clinical images processed with both Hologic and Agfa Musica-2 image processing. Image qualities 3–6 are simulated image qualities from the original clinical images.

			Reader-averaged FoM (95% confidence intervals)		
	Image quality	Image processing	JAFROC	ROC	Reader-averaged LLF at NLF of 0.1
1	Normal dose DR	Hologic	0.83 (0.78, 0.88)	0.91 (0.87, 0.94)	0.70
2	Normal dose DR	Agfa (Musica-2)	0.84 (0.80, 0.88)	0.91 (0.87, 0.95)	0.72
3	Half dose DR	Agfa (Musica-2)	0.68 (0.60, 0.75)	0.82 (0.77, 0.86)	0.47
4	Quarter dose DR	Agfa (Musica-2)	0.52 (0.43, 0.62)	0.70 (0.65, 0.75)	0.27
5	Normal dose CR	Agfa (Musica-2)	0.63 (0.56, 0.70)	0.79 (0.74, 0.84)	0.42
6	Half dose CR	Agfa (Musica-2)	0.55 (0.45, 0.64)	0.70 (0.65, 0.76)	0.30

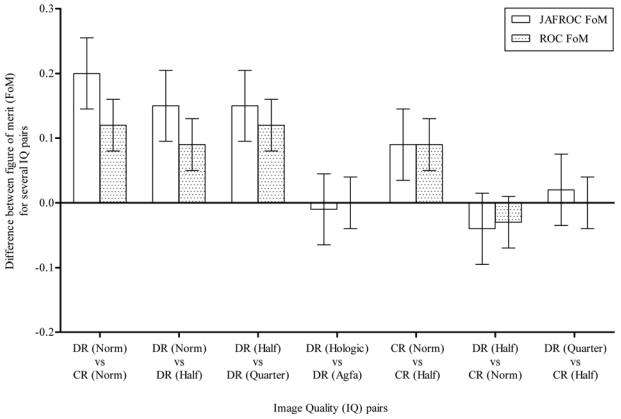

FIG. 8. Difference in FoM between several image quality pairs (error bars indicate 95% confidence intervals).

investigated the effect of dose on detection. Our study also investigated the effect of change in detector providing a wider picture as to the effect on detection of a change in image quality and also allowed us to relate our results to CDMAM measured threshold gold thickness. It is interesting to note that the DR system in the present study happens to

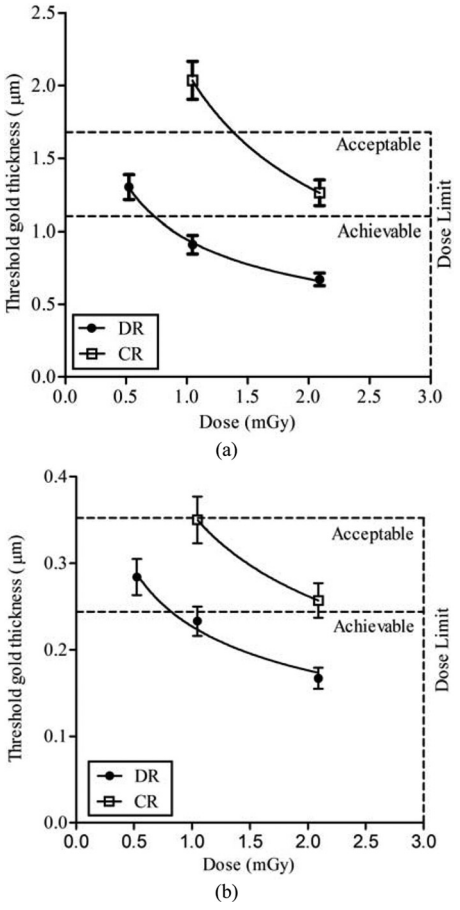

FIG. 9. Threshold gold thickness at five different image qualities: DR at normal, half, and quarter dose levels shown with disc points, and CR at normal and half dose levels shown with square points: (a) 0.1 mm gold disc diameter and (b) 0.25 mm gold disc diameter. Acceptable and achievable standards as set in the European protocol (Ref. 15) are also shown along with dose limit for a breast thickness equivalent to 50 mm PMMA.

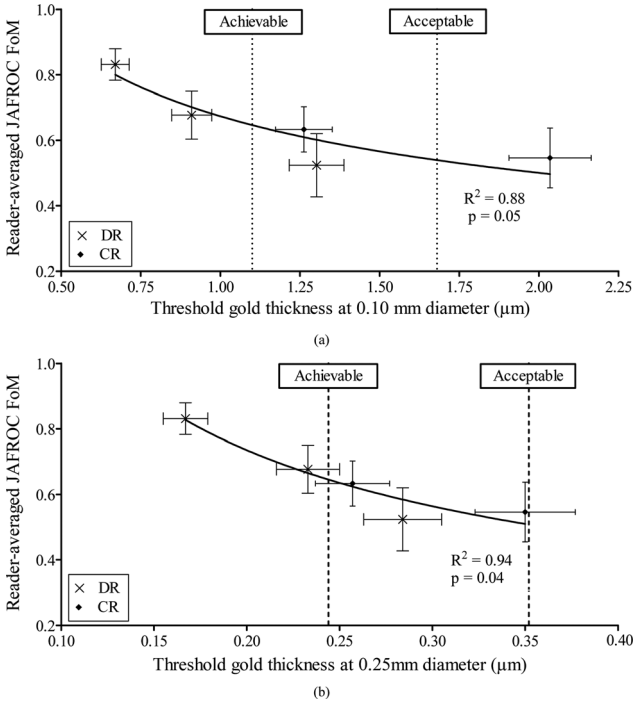

FIG. 10. Reader-averaged JAFROC FoM from the observer study plotted against the threshold gold thickness from CDMAM phantom images at the same IQ for (a) 0.10 mm and (b) 0.25 mm gold disc diameter. DR at normal, half, and quarter dose levels shown with cross points, and CR at normal and half dose levels shown with diamond points The results from the observer study include only images processed using Agfa image processing. The CDMAM analysis was performed on the unprocessed image. Error bars represent the 95% confidence interval. The acceptable and achievable limits as set in the European protocol (Ref. 15) are displayed as dashed lines.

operate at a relatively high dose [average glandular dose (AGD) of 2.09 mGy for breasts 50–60 mm thick]. Oduko *et al.*[42] found that for different DR systems in the United Kingdom the AGD for this breast thickness ranged from 0.8 to 2.2 mGy. Therefore, the half dose level considered here is still clinically relevant.

There was no significant change in microcalcification detection found between the two image processing algorithms investigated which supports a previous smaller study.[12] This result is, however, contrary to the result found by Zanca *et al.*[11] The designs of our study and that of Zanca *et al.* are very similar, suggesting that this difference in results might be due differences in the image processing algorithms investigated. Future research is required to clarify the impact of image processing on calcification and mass detection.

The calcification clusters used in the observer study ranged in diameter and number of calcifications. The diameter of individual calcifications also covered a range of values. Even so, the majority of clusters were small (mean diameter = 3.89 mm), and would be considered subtle. In the screening environment, a much wider range of size and visibility of calcification clusters is present. Calculating the proportion of malignant cancers detected through screening which have the same characteristics as the clusters used in this study would allow us to calculate the expected change in cancer detection in a screening environment which

corresponds to the change in calcification detection found in this study.

We have shown that a good physical performance measured using the CDMAM phantom was matched to good performance in the observer study. This would imply that CDMAM-determined threshold gold thickness is a good predictor of microcalcification detection. All image qualities apart from half dose CR passed the minimum acceptable image quality standard as set in the European protocol,[15] for both the 0.1 and 0.25 mm gold disc diameters (Fig. 9). However, statistical analysis demonstrated significant differences in detection between several image quality pairs (Fig. 8). When considering the optimal use of x-ray imaging technology in breast cancer screening it is important to consider both risk and benefits. These results provide some data from which to estimate the risks and benefits of using greater or lower radiation dose levels. For example, it would seem to be unwise to operate equipment at relatively low dose levels as the reduction in radiation risk may be more than offset by a reduction in cancer detection. Similarly, these results suggest that the use of better quality detectors may improve cancer detection at the same dose levels. Such considerations are likely to lead to a revision of the standards in the European Guidelines to ensure adequate detection of calcifications. One option would be to require that systems be as good as or better than the achievable image quality level to optimize calcification detection while meeting existing dose limits. It is expected that most modern DR system could meet such a standard.

Our study has two main limitations. Only calcification clusters were inserted and not other radiological features such as masses. The relationship between image quality and detection may differ for various radiological features, and so investigating both microcalcification and mass detection is important. Also the calcifications inserted were all malignant. Introducing benign calcification clusters would allow any difference in recall (i.e., the interpretation of the feature as well as detection) between the different image qualities to be assessed.

V. CONCLUSIONS

Significant differences were found between detection of subtle calcification clusters in CR and DR images at the same dose level. There was also a significant reduction in detection with reduced dose for both CR and DR images. There was no significant difference in detection between the two image processing algorithms investigated.

When relating the results of the observer study to the measured threshold gold thicknesses for 0.25 and 0.1 mm gold disc diameters, a smaller threshold gold thickness correlated with better performance in the observer study. This is an important new finding and demonstrates that threshold gold thickness measurements using a CDMAM phantom relate to calcification detection. However, when relating measured threshold gold thickness measurements to European standards for mammographic image quality, image qualities with significantly poorer calcification detection rates still gave better performance than the current minimum acceptable standard. This suggests that the current EU standards may need revising.

ACKNOWLEDGMENTS

This work was part of the OPTIMAM project and was supported by Cancer Research, UK, and the Engineering and Physical Sciences Research Council Cancer Imaging Programme in Surrey, in association with the Medical Research Council and Department of Health (England). One of the authors (DPC) was supported in part by grants from the Department of Health and Human Services, National Institute of Health, ROI-EB005243 and ROI-EB008688. The authors are grateful to Anna Cummin, Carole Kliger, Caroline Taylor, Karen Mitchell, Susan Milton, Victoria Cooke, and Vanessa Newman for taking part in the study. The authors would also like to thank the collaborators from Katholieke Universiteit Leuven in particular Federica Zanca for the advice on study design and JAFROC statistical analysis, and Jurgen Jacobs for providing the SARA software. The authors are thankful for the help and co-operation of staff from the Departments of Radiology, Breast Screening, and Pathology at St. Georges Hospital (London), in particular Susan Clarke, Sue Bailley, Geoff Cattini, Ruth Nash, and Val Thomas. For their assistance during measurements, the authors are also thankful to Jenny Oduko, Lebina Shrestha, and Faith Green at NCCPM. Finally, the authors thank the companies Hologic, Inc., Carestream Healthcare, Agfa Healthcare, and MIS Healthcare for their cooperation.

[a]Electronic mail: lucy.warren@nhs.net

[1]E. D. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J. Baum, S. Acharyya, E. F. Conant, L. L. Fajardo, L. Bassett, C. D'Orsi, R. Jong, and M. Rebner, "Diagnostic performance of digital versus film mammography for breast-cancer screening," N. Engl. J. Med. **353**, 1773–1784 (2005).

[2]P. Skaane and A. Skjennald, "Screen-film mammography versus full-field digital mammography with soft-copy reading: Randomized trial in a population-based screening program—The Oslo II study," Radiology **232**, 197–204 (2004).

[3]R. E. Hendrick, E. B. Cole, E. D. Pisano, S. Acharyya, H. Marques, M. A. Cohen, R. A. Jong, G. E. Mawdsley, K. M. Kanal, C. J. D'Orsu, M. Rebner, and C. Gatsonis, "Accuracy of soft-copy digital mammography versus that of screen-film mammography according to digital manufacturer: ACRIN DMIT reterospective multireader study," Radiology **247**, 38–48 (2008).

[4]X. J. Rong, C. C Shaw, S. K. Thompson, K. T. Krugh, C. Lai, D. A. Johnston, M. R. Lemocks, X. Liu, G. J. Whitman, M. J. Dryden, and T. W. Stephens, "Micro-calcification detectability for four mammographic detectors: Flat-panel, CCD, CR and screen/film," Med. Phys. **29**, 2052–2062 (2002).

[5]G. Schueller, C. Riedl, R. Mallek, K. Eibenberger, H. Langenberger, E. Kaindl, C. Kulinna-Cosentini, M. Rudas, and T. Helbich, "Image quality, lesion detection and diagnostic efficacy in digital mammography: Full-field digital mammography versus computed radiography-based mammography using digital storage phosphor plates," Eur. J. Radiol. **67**, 487–496 (2008).

[6]M. Ruschin, P. Timberg, M. Bäth, B. Hemdal, T. Svahn, R. Saunders, E. Samei, I. Andersson, S. Mattsson, D. P. Chakraborty, and A. Tingberg, "Dose dependence of mass and microcalcification detection in digital mammography: Free response human observer studies," Med. Phys. **34**, 400–407 (2007).

[7]M. Yakabe, S. Sakai, H. Yabuuchi, Y. Matsuo, K. Takeshi, T. Setoguchi, M. Cho, M. Masuda, and M. Sasaki, "Effect of dose reduction on the

ability of digital mammography to detect simulated micro-calcifications," J. Digit Imaging **23**, 520–526 (2010).
[8]E. Samei, R. Saunders, J. Baker, and D. Delong, "Digital Mammography: Effects of reduced radiation dose on diagnostic performance," Radiology **243**, 396–404 (2007).
[9]D. P. Chakraborty, "New developments in observer performance methodology in medical imaging," Semin. Nucl. Med. **41**, 401–418 (2011).
[10]E. B. Cole, E. D. Pisano, D. Zeng, K. E. Muller, S. R. Aylward, S. Park, C. Kuzmiak, M. Koomen, D. Pavic, R. Walsh, J. Baker, E. I. Gimenez, R. Freimanis, and K. Muller, "The effects of gray scale image processing on digital mammography interpretation performance," Acad. Radiol. **12**, 585–595 (2005).
[11]F. Zanca, J. Jacobs, C. V. Ongeval, F. Claus, V. Celis, C. Geniets, V. Provost, H. Pauwels, G. Marchal, and H. Bosmans, "Evaluation of clinical image processing algorithms used in digital mammography," Med. Phys. **36**, 765–775 (2009).
[12]T. Uematsu, "Detection of masses and calcifications by soft-copy reading: Comparison of two post-processing algorithms for full-field digital mammography," Jpn. J. Radiol. **27**, 168–175 (2009).
[13]E. B. Cole, E. D. Pisano, E. D. Kistner, K. E. Muller, M. E. Brown, S. A. Feig, R. A. Jong, A. D. A. Maidment, M. J. Staiger, C. M. Kuzmiak, R. I. Freimanis, N. Lesko, E. L. Rosen, R. Walsh, M. Williford, and P. Braeuming, "Diagnostic accuracy of digital mammography in patients with dense breasts who underwent problem-solving mammography: Effects of image processing and lesion type," Radiology **226**, 153–160 (2002).
[14]A. Mackenzie, A. Workman, D. R. Dance, M. Yip, K. Wells, and K. C. Young, "Conversion of mammographic images to appear with noise and sharpness characteristics of a different detector and x-ray system," Med. Phys. **39**, 2721–2734 (2012).
[15]European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis, *The European Protocol for the Quality Control of the Physical and Technical Aspects of Mammography Screening, Part B: Digital Mammography*, 4th ed. (European Commission, Luxembourg, 2006).
[16]A. K. Carton, H. Bosmans, C. Vanongeval, G. Souverijns, G. Marchal, J. Jacobs, D. Vandenbroucke, H. Pauwels, and K. Nijs, "Contrast threshold of 4 full field digital mammography systems using different measurement methods," in *IWDM*, LNCS (Springer-Verlag, Berlin, Heidelberg, 2006), Vol. 4046, pp. 593–600.
[17]F. Zanca, D. P. Chakraborty, C. V. Ongeval, J. Jacobs, F. Claus, G. Marchal, and H. Bosmans, "An improved method for simulating microcalcifications in digital mammograms," Med. Phys. **35**, 4012–4018 (2008).
[18]A. K. Carton, H. Bosmans, C. Van Ongeval, G. Souverijns, F. Rogge, A. Van Steen, and G. Marchal, "Development and validation of a simulation procedure to study the visibility of micro-calcifications in digital mammograms," Med. Phys. **30**, 2234–2241 (2003).
[19]S. L. Hillis, N. A. Obuchowski, and K. S. Berbaum, "Power estimation for multireader ROC methods: An updated and unified approach," Acad. Radiol. **18**, 129–142 (2011).
[20]A. K. Carton, H. Bosmans, D. Vandenbroucke, G. Souverjins, C. V. Ongeval, O. Dragusin, and G. Marchal, "Quantification of Al-equivalent thickness of just visible micro-calcifications in full field digital mammograms," Med. Phys. **31**, 2165–2176 (2004).
[21]F. Zanca, C. V. Ongeval, N. Marshall, T. Meylaers, K. Michielsen, G. Marchal, and H. Bosmans, "The relationship between the attenuation properties of breast micro-calcifications and aluminum," Phys. Med. Biol. **55**, 1057–1068 (2010).
[22]D. R. Dance, C. L. Skinner, K. C. Young, and R. E. van Engen, "Additional factors for the estimation of mean glandular breast dose using the UK mammography dosimetry protocol," Phys. Med. Biol. **45**, 3225–3240 (2000).
[23]D. R. Dance and G. J. Day, "The computation of scatter in mammography by Monte Carlo methods," Phys. Med. Biol. **29**, 237–247 (1984).
[24]J. M. Boone, K. K. Lindfors, V. N. Cooper III, and J. A. Seibert, "Scatter/primary in mammography: Comprehensive results," Med. Phys. **27**, 2408–2416 (2000).
[25]I. Sechopoulos, S. Suryanarayanan, S. Vedantham, C. J. D'Orsi, and A. Karellas, "Scatter radiation in digital tomosynthesis of the breast," Med. Phys. **34**, 564–576 (2007).
[26]F. Zanca, G. Zhang, N. Marshall, E. Shaheen, E. Salvagnini, G. Marchal, and H. Bosmans, "Software framework for simulating clusters of microcalcifications in digital mammography," in *IWDM*, LNCS (Springer-Verlag, Berlin, Heidelberg, 2010), Vol. 6136, 689–696.
[27]M. Yip, A. Mackenzie, E. Lewis, D. R. Dance, K. C. Young, W. Christmas, and K. Wells, "Image resampling in mammographic image simulation," Phys. Med. Biol. **56**, N275–N286 (2011).
[28]J. Jacobs, F. Zanca, and H. Bosmans, "A novel platform to simplify human observer performance experiments in clinical reading environments," Proc. SPIE **7966**, 79660B (2011).
[29]NHS Breast Screening Programmes, *Quality Assurance Guidelines for Breast Cancer Screening Radiology*, 2nd ed., edited by J. Liston and R. Wilson (Sheffield, England, 2011), Publication No. 59.
[30]E. Samei *et al.*, "Assessment of display performance for medical imaging systems: Executive summary of AAPM TG18 report," Med. Phys. **32**, 1205–1226 (2005).
[31]J. Wei, H. P. Chan, C. Zhou, Y. T. Wu, B. Sahiner, L. M. Hadjiiski, M. A. Roubidoux, and M. A. Helvie, "Computer-aided detection of breast masses: Four-view strategy for screening mammography," Med. Phys. **38**, 1867–1876 (2011).
[32]F. Zanca, C. V. Ongeval, J. Jacobs, G. Marchal, and H. Bosmans, "A quantitative method for evaluating the detectability of lesions in digital mammography," Radiat. Prot. Dosim. **129**, 214–218 (2008).
[33]D. P. Chakraborty and T. Svahn, "Estimating the parameters of a model of visual search from ROC data: An alternative method for fitting proper ROC curves," Proc. SPIE **7966**, 79660L1 (2011).
[34]D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: Modelling, analysis, and validation," Med. Phys. **31**, 2313–2330 (2004).
[35]D. D. Dorfman, K. S. Berbaum, and C. E. Metz, "Receiver operating characteristic rating analysis: Generalization to the population of readers and patients with jackknife method," Invest. Radiol. **27**, 723–731 (1992).
[36]B. Zheng, J. K. Leader, G. Abrams, B. Shindel, V. Catullo, W. F. Good, and D. Gur, "Computer-aided detection schemes: The effect of limiting the number of cued regions in each case," AJR **182**, 579–583 (2004).
[37]H. P. Chan, J. Wei, Y. Zheng, M. Helvie, R. H. Moore, B. Sahiner, L. Hadjiiski, and D. B. Kopans, "Computer-aided detection of masses in digital tomosynthesis mammography: Comparison of three approaches," Med. Phys. **35**, 4087–4095 (2008).
[38]K. C. Young, J. J. H. Cook, and J. M. Oduko, "Automated and human determination of threshold contrast for digital mammography systems," in *IWDM*, LNCS (Springer-Verlag, Berlin, Heidelberg, 2006), Vol. 4046, pp. 266–272.
[39]K. C. Young, A. Alsager, J. M Oduko, H. Bosmans, B. Verbrugge, T. Geertse, and R. van Engen, "Evaluation of software for reading images of the CDMAM test object to assess digital mammography systems," Proc. SPIE **6913**, 69131C (2008).
[40]A. Workman, I. Castellano, E Kulama, C. P. Lawinski, N. Marshall, and K. C. Young, "Commissioning and routine testing of full field digital mammography systems," NHSBSP Equipment Report No. 0604, Version 2 (2006).
[41]N. W. Marshall, K. Lemmens, and H. Bosmans, "Physical evaluation of a needle photostimulable phosphor based CR mammography system," Med. Phys. **39**, 811–824 (2012).
[42]J. M. Oduko, K. C. Young, and A. Burch, "A survey of patient doses from digital mammography systems in the UK in 2007 to 2009," in *Digital Mammography*, Lecture Notes in Computer Science (Springer-Verlag, Berlin, Heidelberg, 2010), Vol. 6136, pp. 365–370.